Title: Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: The PRISMA-DTA statement

Subtitle: The PRISMA-DTA statement.

Author List:

1. Matthew DF McInnes MD. The University of Ottawa Department of Radiology. The Ottawa Hospital Research Institute Clinical Epidemiology Program.
2. David Moher PhD. The Ottawa Hospital Research Institute Clinical Epidemiology Program (Centre for Journalology). dmoher@ohri.ca
3. Brett D Thombs PhD, Lady Davis Institute of the Jewish General Hospital and Department of Psychiatry, McGill University. brett.thombs@mcgill.ca
4. Trevor A McGrath BSc. The University of Ottawa Department of Radiology. tmcgr043@uottawa.ca
5. Patrick M Bossuyt PhD, Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Academic Medical Center, Amsterdam. p.m.bossuyt@amc.nl

Additional Authors: the PRISMA-DTA Group (PRISMA-DTA Group member authors are listed on the following page)

Corresponding Author: Matthew McInnes, Rm c-159 Department of Medical Imaging, The Ottawa Hospital- Civic Campus, 1053 Carling Ave, Ottawa, ON, K1E 4Y9, 613-761-4054, fax 613-761-4476. Email: mmcinnes@toh.ca

Date of Revision: November 30, 2017

Word Count (introduction to discussion): 2937 words

Author List: The PRISMA-DTA Group

Name	Email	Affiliation(s)
Tammy Clifford, PhD	tammyc@cadth.ca	Canadian Agency for Drugs and Technologies in Health, Ottawa, Canada
Jérémie F Cohen, MD, PhD	jeremie.cohen@inserm.fr	1. Department of Pediatrics, Necker - Enfants Malades Hospital, Assistance Publique - Hôpitaux de Paris, Paris Descartes University, France. 2. Inserm UMR 1153, Research Center for Epidemiology and Biostatistics Sorbonne Paris Cité (CRESS), Paris Descartes University, France
Jonathan J. Deeks, PhD	j.deeks@bham.ac.uk	University of Birmingham, UK
Constantine Gatsonis, PhD	gatsonis@stat.brown.edu	Brown University, USA
Lotty Hooft, PhD	l.hooft@umcutrecht.nl	Cochrane Netherlands, Julius Center for Health Sciences and Primary Care, UMC Utrecht
Harriet A. Hunt, MSc	H.A.Hunt@exeter.ac.uk	University of Exeter, UK
Christopher J. Hyde, PhD	C.J.Hyde@exeter.ac.uk	University of Exeter
Daniël A. Korevaar, MD, PhD	d.a.korevaar@amc.uva.nl	AMC Amsterdam, Netherlands
Mariska M.G. Leeflang, PhD	m.m.leeflang@amc.uva.nl	AMC Amsterdam, Netherlands
Petra Macaskill, PhD	petra.macaskill@sydney.edu.au	University of Sydney, Australia
Johannes B. Reitsma, MD, PhD	j.b.reitsma-2@umcutrecht.nl	Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Cochrane Netherlands
Rachel Rodin, MD, MPH	rachel.rodin@phac-aspc.gc.ca	Public Health Agency of Canada, Ottawa, Canada
Anne W.S. Rutjes, PhD	Anne.rutjes@ispm.unibe.ch	ISPM & BIHAM, University of Bern, Switzerland
Jean-Paul Salameh BSc	jsala016@uottawa.ca	University of Ottawa, School of Epidemiology and Public Health, The Ottawa Hospital Research Institute Clinical Epidemiology Program
Adrienne Stevens, MSc	adstevens@ohri.ca	Ottawa Hospital Research Institute; Translational Research in Biomedicine (TRIBE) Program, University of Split School of Medicine, Split, Croatia
Yemisi Takwoingi, PhD	y.takwoingi@bham.ac.uk	University of Birmingham, UK
Marcello Tonelli, MD, SM	cello@ucalgary.ca	University of Calgary, Canada
Laura Weeks, PhD	LauraW@cadth.ca	Canadian Agency for Drugs and Technologies in Health, Ottawa, Canada
Penny Whiting, PhD	penny.whiting@bristol.ac.uk	University of Bristol & NIHR CLAHRC West, UK
Brian H. Willis, MD, PhD	b.h.willis@bham.ac.uk	University of Birmingham, UK

Key Points

Question: What items should be reported to allow readers to evaluate the validity and applicability and to enhance the replicability of diagnostic test accuracy (DTA) systematic reviews?

Findings: This guideline is an extension of the PReferred Items for reporting Systematic Reviews and Meta-Analyses (PRISMA): PRISMA-DTA. Two PRISMA items have been omitted, two added, and 17 were modified to reflect DTA-specific or optimal contemporary systematic review methods.

Meaning: PRISMA-DTA can facilitate transparent reporting of DTA reviews, and may help assist evaluations of validity and applicability, enhance replicability of reviews, and make the results more useful for clinicians, journal editors, reviewers, guidelines-authors and funders.

Abstract

IMPORTANCE: Diagnostic test accuracy (DTA) systematic reviews synthesize data from primary diagnostic studies that have evaluated the accuracy of one or more index tests against a reference standard. DTA systematic reviews provide estimates of test performance, allow comparisons of the accuracy of different tests, and facilitate the identification of sources of variability in test accuracy.

OBJECTIVE: To develop PRISMA-DTA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) as a stand-alone extension of the PRISMA Statement, modified to reflect the particular requirements for reporting DTA systematic reviews and meta-analyses, as well as PRISMA-DTA for abstracts.

DESIGN: Established standards for guideline development (EQUATOR Network) were followed. The PRISMA-Statement was used as a framework upon which to modify and add material. A group of 24 multi-disciplinary experts used a systematic review of existing reporting guidelines and methods articles, a 3-round Delphi process, a consensus meeting, pilot testing, and iterative refinement to develop PRISMA-DTA. The final version of the PRISMA-DTA checklist was approved by the PRISMA-DTA group.

FINDINGS: Systematic review (64 items) and Delphi feedback (six new items proposed and one item split into two) identified 71 potentially relevant items for consideration; the Delphi process reduced these to 60 items that were discussed at the consensus meeting. Following the meeting, piloting and iterative feedback was used to generate the 27-item PRISMA-DTA checklist. To reflect DTA-specific or optimal contemporary systematic review methods, 8 of the 27 original PRISMA items were left unchanged, 17 were modified, 2 were added and 2 omitted.

CONCLUSIONS AND RELEVANCE: PRISMA-DTA provides specific guidance for reporting of DTA systematic reviews. PRISMA-DTA can facilitate transparent reporting of DTA reviews, and may assist evaluations of validity and applicability, enhance replicability of reviews, and make the results more useful for multiple stakeholders.

Introduction

Systematic reviews can advance understanding of diagnostic test accuracy (DTA). DTA systematic reviews synthesize data from primary studies to provide insight into the ability of medical tests to detect a target condition—they can provide estimates of test performance, allow comparisons of the accuracy of different tests, and facilitate the identification of sources of variability[1]. The number of DTA systematic reviews has increased rapidly; however, they are often not reported completely, which has contributed to "a crisis of repeatability"[2-5].

Reporting of systematic reviews should be complete and informative to enable readers to assess the quality of methods and validity of findings. Published DTA systematic reviews are often not informative, and are of heterogeneous quality[4,6,7]. They demonstrate variability in approaches to fundamental methodological steps, including methods to assess risk of bias, assessment of between-study variability, and methods for combining data across studies[7-11].

To improve reporting of systematic reviews, the Preferred Reporting Items for Systematic-Reviews and Meta-Analyses (PRISMA) guideline was published (27-item checklist and flow-diagram)[13]. PRISMA was directed at systematic reviews of interventions; the authors suggested modification for DTA reviews [14]. Though DTA systematic-reviews share elements with those of interventions, there are important differences. Study-design and measures of effect differ from those of randomized-trials. Accuracy can differ between studies due to differences in patients, setting, prior testing and use of different reference standards. Consequently, methods for evaluating risk of bias, summarizing results, and exploring variability differ from those used for interventions. As such, some PRISMA-items are not appropriate for DTA systematic reviews, others need adaptation, and some areas may not be covered[1,15,16].

We aimed to develop PRISMA-DTA as a stand-alone extension of PRISMA, modified to reflect the particular requirements for reporting DTA systematic reviews. A secondary objective was to identify items that should be included in abstracts of DTA systematic reviews (PRISMA-DTA for Abstracts).

Methods

This study was determined to be exempt from IRB-review. After establishing the PRISMA-DTA executive group, comprised of the lead author of PRISMA (DM) [13], the lead author of STARD (Standards for the Reporting of Diagnostic Accuracy Studies) (PMMB) [17] and an experienced author, reviewer and editor of DTA systematic reviews (MDFM), a number of experts were contacted to join the PRISMA-DTA group and assist with the project (all contacted experts agreed to participate). The goal was to assemble a team of experts in DTA research and systematic review methods, complemented by DTA systematic review authors, journal editors, funders, and DTA systematic review users. eTable 1 lists the 24 members and their relevant expertise of the PRISMA-DTA group.

The PRISMA-DTA executive registered the protocol for developing the statement with the Enhancing the QUAlity and Transparency Of health Research (EQUATOR) network [19]. The protocol was based on previously published guidance for establishing reporting guidelines, created by the EQUATOR network; no major deviations from the protocol occurred [20]. The PRISMA-DTA group used the standard PRISMA statement [13] as a starting point and endeavored to identify items that needed to be added, removed or modified to meet the needs of DTA systematic reviews.

Details of the systematic review for item generation have been published elsewhere[21]. Briefly, searches of multiple databases and existing sources of guidance for reporting of systematic reviews and DTA studies (e.g. PRISMA, STARD 2015) [13,17] were performed to identify articles pertaining to methodology or reporting quality of DTA systematic reviews. Following data extraction from these reports, potential PRISMA-DTA items were categorized according to specific reporting topics: general overview, quality of reporting, search, variability, pooling methods, publication bias, risk of bias, and 'other'. This list of potential PRISMA-DTA items was presented in the first round of the Delphi procedure.

Delphi Procedure

A three-round Delphi process was held between December 2016-March 2017, in which all members of the PRISMA-DTA group were invited to participate [22,23]. This modified Delphi process has been used previously for similar work such as: Risk of Bias in Systematic reviews (ROBIS) and STARD 2015 [24,25]. The aim of the process was to

achieve consensus on essential items that should make up PRISMA-DTA and to identify items that required discussion at the consensus meeting.

During each round of the survey, potential essential items were proposed, and participants were asked to score each item on a 1-5 Likert scale, anchored at (1) "not essential to report in a systematic review of diagnostic test accuracy studies" and (5) "essential to report in a systematic review of diagnostic test accuracy studies". Likert scores were categorized as: 1-2 = low score (item should not be part of PRISMA-DTA), 3 = moderate (item should be discussed), 4-5 = high score (item should be part of PRISMA-DTA). For an item to meet 'consensus', more than 66% of the Delphi respondents needed to rate one of these three categories; this threshold is based on that used for previous reporting guidelines[26].

During round 1 of the Delphi survey, all items identified during the systematic review step were proposed[21]. Participants were also asked to suggest any additional items that were potentially relevant to report in systematic reviews of DTA studies. Round 2 of the survey included any items that did not reach consensus in round 1, and any new items suggested by at least 1 respondent in round 1. As with round 2, round 3 involved items that did not reach consensus in rounds 1 or 2.

Following each of the three rounds, the mode (most frequent) score for each item was tabulated. Items were categorized as follows: a) mode score 1-3 but <66% of participants: proceed to next round of Delphi (or to meeting discussion if Delphi round 3), b) consensus score 1 or 2: do not include; c) consensus score 3: discuss at meeting; d) mode score 4 or 5 but for <66% of participants: discuss at meeting; e) consensus score 4 or 5: include in PRISMA-DTA (but discuss at meeting to confirm exact wording). All participants were provided an anonymized summary of the results after each round of the process. SurveyMonkey (SurveyMonkey Inc, San Mateo, California, USA) was used to administer the survey.

Consensus meeting

A two-day consensus meeting was held in Amsterdam, The Netherlands in May, 2017; all members of the executive and PRISMA-DTA group were invited to attend. The main objective of this meeting was to agree on items for which no consensus was reached

during the Delphi survey, to generate a preliminary PRISMA-DTA checklist (and PRISMA-DTA for abstracts). For the items that reached consensus for inclusion prior to the meeting, precise wording of items was decided.

Checklist Pilot

Following the meeting, members of the PRISMA-DTA group reviewed and applied the checklist to ongoing DTA systematic reviews in order to identify any practical challenges with any of the items and to inform writing of the statement. This included formal piloting of the preliminary checklist on published DTA systematic reviews by a graduate student (JPS). In addition, multiple potential users and stakeholders were invited to review and apply the preliminary checklist in order to assess utility and clarity of wording. Feedback from these pilot exercises was used to refine wording and presentation of the checklist. Formal feedback was gathered via a survey administered via SurveyMonkey (SurveyMonkey Inc, San Mateo, California, USA); this was sent to the entire PRISMA-DTA group. Additional feedback was gathered via email correspondence. All sources of feedback were used to modify and inform the final version of the PRISMA-DTA checklist.

A further explanation and elaboration document (E&E) will subsequently be developed to provide additional detail regarding the rationale for items and examples.

Results

Delphi procedure

Participation in the Delphi survey is documented in eTable 1; 23 individuals completed all 3 rounds of the survey. In Round 1, the group evaluated 64 items identified by the systematic review: 42 items met consensus for inclusion, 20 items moved forward to round 2, 2 items were excluded, and an additional 6 items were suggested for inclusion for round 2. In Round 2, the group assessed 27 items (1 item from round 1 was split into 2): 5 items met consensus for inclusion, 15 items moved forward to round 3, and 7 items were excluded. In Round 3, no item met consensus for inclusion, 13 items were moved forward to the consensus meeting, and 2 items were excluded. Overall, after 3 Delphi rounds, 47 items were included (final wording to be discussed at the face-to-face

consensus meeting), 13 items moved forward to the consensus meeting to discuss inclusion or exclusion, and 11 items were excluded. A list of the 11 excluded items is provided in eTable 2. While these items are considered relevant to reporting of DTA systematic reviews, they were felt to be either too detailed for a 'minimum reporting guideline' or not relevant depending on the scope or purpose of the review. Several of these items will be further discussed in the forthcoming Explanation and Elaboration document. Flow diagram documenting the Delphi procedure is provided in Figure 1.

Consensus Meeting

Meeting attendance (n = 18) and the agenda are documented in eTables 1 and 3 respectively. Of the 60 items discussed at the meeting, 27 were excluded. Excluded items and the rationale for exclusion are provided in eTable 2.

Two of the 27 items of the original PRISMA checklist were confirmed for removal: items 15 and 22. These items refer to evaluation for and reporting of risk of bias that may affect the cumulative evidence such as publication bias and selective reporting within studies. These were excluded for two main reasons. First, there is only limited evidence that publication or reporting bias is a major issue for primary DTA studies[27,28]. As such, the rationale for mandating its evaluation in DTA systematic reviews is not as strong as for intervention reviews. Second, there is no appropriate test with adequate statistical power to reliably assess publication bias in the context of DTA systematic reviews [29-31].

The remaining 33 items were discussed and synthesized into a draft PRISMA-DTA checklist; many of the items were combined in order to reduce redundancy between and minimize the total number of items. The PRISMA flow diagram was also reviewed at the consensus meeting and no modifications for PRISMA-DTA were deemed necessary.

Compared to the original PRISMA checklist, two new items were added. Item D1: State the scientific and clinical background, including the intended use and clinical role of the index test, and if applicable, the rationale for minimally acceptable test accuracy (or minimum difference in accuracy for comparative reviews). The rationale for inclusion is two-fold. First, the role of the index test is critical to understanding the place

of a test in the diagnostic pathway; diagnostic accuracy can vary importantly depending on the clinical scenario. Without this information, generalizability of the results to the clinical setting may be limited [17,18]. Second, identifying minimum acceptable test accuracy may be helpful in forming conclusions. Whether a test is considered clinically useful cannot be determined by a diagnostic accuracy measure alone; its accuracy relative to alternative tests or management strategies must be considered, as well as the downstream consequences of false positives and false negatives. As such, considering external evidence to form criteria for minimum acceptable test standards may play an important role in forming the purpose of DTA systematic reviews [12,17,18].

Defining a minimally acceptable test accuracy (or minimum difference) may not always be appropriate depending on the review question. For example, if a test is not yet well established or understood, the purpose of the review might be to evaluate reasons for variability in accuracy. For this reason, we have added the qualifier 'if applicable' to this item.

The second new item was D2: Report the statistical methods used for meta-analyses, if performed. Meta-analyses of DTA studies typically require multivariate models (e.g. bivariate and hierarchical summary ROC) which allow for the trade-off between sensitivity and specificity due to positivity threshold, for potential correlation between estimates of sensitivity and specificity across studies and for variability through the inclusion of random effects [32,33]. Traditional univariate methods ignore this correlation and can give misleading results [5,34,35]. We acknowledge that there are instances when univariate methods may be appropriate, for example if the specificity of a test is set at 100%, and univariate meta-analysis of sensitivity is the focus of the review. As such, reporting the method used for meta-analysis (if done) was considered essential for DTA systematic reviews.

Eight of the original PRISMA items were not modified since they were felt to be equally applicable to systematic reviews of DTA (Original PRISMA items: 3, 5, 7, 9, 10, 16, 17 and 27).

Seventeen of the original PRISMA items were adapted: (items 1, 2, 4, 6, 8, 11-14, 18-21, and 23-26). The reason for modification varied—the two major reasons were either that there was unclear or ambiguous wording in the original PRISMA statement

that required updating or that wording specific to issues for DTA systematic reviews was necessary. Table 1 lists the rationale for modification of the original PRISMA items for PRISMA-DTA. Further explanation and elaboration on the rationale and evidence will be provided in the forthcoming Explanation & Elaboration document.

At the consensus meeting, PRISMA for Abstracts was modified to arrive at PRISMA-DTA for Abstracts [36]. The total number of items (n=12) was preserved. Five items were not modified (Items 4, 6, 10-12). One item was deleted (Item 8- description of the effect) since effect size is only relevant to interventions, and not DTA studies [1,29]. One new item was added (A1. Synthesis of results); this corresponds to new item D2 in PRISMA-DTA. Six items were modified (Items 1-3, 5, 7 and 9) to reflect the modified language for the corresponding items in PRISMA-DTA (as described in Table 1 and above).

Piloting & Revision

Thirty-seven points of feedback from the pilot exercise were received via email and formal survey. This feedback was considered by the PRISMA-DTA executive group and used to modify five of the items, as well as to add further explanation and rationale.

Final Checklists

The final version of the PRISMA-DTA checklist is provided in Table 2. PRISMA-DTA has the same number of items as PRISMA; 2 original items were deleted (items 15 and 22) and 2 new items were added. Numbering from the original PRISMA statement is preserved; the 2 new items are labelled: D1 and D2.

The final version of the PRISMA-DTA for Abstracts checklist is provided in Table 3. PRISMA-DTA for abstracts has the same number of items as PRISMA for abstracts (n=12) One item was deleted (item 8) and one new item was added. Numbering from the original PRISMA for abstracts is preserved; the new item is labelled: A1.

Discussion

PRISMA-DTA is a reporting guideline providing guidance specific to DTA systematic reviews. Both the PRISMA-DTA statement and PRISMA-DTA for Abstracts were

developed with multidisciplinary consensus approaches as per best practices for guideline development [20]. PRISMA-DTA items reflect the concepts, methodology and language specific to DTA systematic reviews and, if implemented, can help ensure that information for assessment of risk of bias and applicability in DTA research is reported, and can enhance transparency and replicability of DTA systematic reviews. This work should be of practical use to multiple stakeholders, who author, review, publish, fund and implement results of DTA systematic reviews; it may also be useful as a guidance for DTA systematic review protocols. DTA systematic reviews for which this checklist is relevant include evaluation of single tests, multiple tests (comparative), and multivariable diagnostic models.

PRISMA-DTA aims to improve completeness and transparency of reporting of DTA systematic reviews. Complete reporting might be associated with review quality, however, they are not inseparable [4]. Understanding and application of optimal DTA systematic review principles and methods is fraught with complexity and pitfalls—it cannot all be learned from a 27-item reporting checklist [1]. While guidance is available for conducting DTA systematic reviews [29], considerable areas of uncertainty remain (e.g. optimal methods for assessing variability, appropriate interpretation of review findings); these areas are likely to evolve based on ongoing and future research [8]. As such, prospective reviewers are encouraged to seek specialized training (e.g. Cochrane Screening and Diagnostic Tests Methods Group author training resources) and to collaborate with those experienced in DTA systematic review methods[37].

Conforming to reporting guidelines can be challenging based on journal-level constraints such as limits on words, tables and figures although there is little evidence to indicate that reporting guidelines increase article word length. Methods to ensure complete reporting may include the use of supplementary on-line only material, institutional repositories and appendices. PRISMA-DTA represents minimum reporting requirements, rather than a constraint or cap on what should be reported. Additional information that authors consider relevant to their specific review question may also be reported (e.g. inter-observer agreement for imaging reviews). Complete reporting of DTA systematic reviews may be hindered by incomplete reporting in DTA primary studies[38]. This challenge makes complete reporting of DTA systematic reviews all the

more important since readers need to know whether necessary information from the primary studies was available from which conclusions can be drawn.

Limitations

The PRISMA-DTA project was guided by evidence-based principles where possible; however, when evidence was lacking, expert opinion was relied upon. PRISMA-DTA was designed for all types of DTA research; some specialties (e.g. imaging) may have important items unique to their specialty (e.g. inter-observer agreement) that are not included in PRISMA-DTA and should be reported. Finally, as the body of evidence in DTA research grows, PRISMA-DTA will need to be updated to reflect these advances.

Conclusion

PRISMA-DTA provides specific guidance for reporting of DTA systematic reviews. PRISMA-DTA can facilitate transparent reporting of DTA reviews, and may assist evaluations of validity and applicability, enhance replicability of reviews, and make the results more useful for multiple stakeholders.

Note to reviewers/ editors: *the text box below is a suggested stand-alone summary of changes to be included in the paper.*

FROM PRISMA TO PRISMA-DTA: HOW DID PRISMA ITEMS CHANGE?

PRISMA ITEMS REMOVED (2)

Item 15: Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).
Item 22: Present results of any assessment of risk of bias across studies.

PRISMA ITEMS MODIFIED (17)

Items 1, 2, 4, 6, 8, 11-14, 18-21 and 23-26.

NEW ITEMS (2)

D1: State the scientific and clinical background, including the intended use and clinical role of the index test, and if applicable rationale for minimally acceptable test accuracy (or minimum difference in accuracy for comparative design).
D2: Report the statistical methods used for meta-analyses, if performed.

Details and rationale for removal, modifications and additions are provided in the text and appendices.

References

1. McInnes MD, Bossuyt PM. Pitfalls of Systematic Reviews and Meta-Analyses in Imaging Research. *Radiology*. 2015;277(1):13-21.
2. Bastian H, Glasziou P, Chalmers I. Seventy-five trials and eleven systematic reviews a day: how will we ever keep up? *PLoS Med*. 2010;7(9):e1000326.
3. Glasziou P, Altman DG, Bossuyt P, et al. Reducing waste from incomplete or unusable reports of biomedical research. *Lancet*. 2014;383(9913):267-276.
4. Tunis AS, McInnes MD, Hanna R, Esmail K. Association of study quality with completeness of reporting: have completeness of reporting and quality of systematic reviews and meta-analyses in major radiology journals changed since publication of the PRISMA statement? *Radiology*. 2013;269(2):413-426.
5. McGrath T, McInnes M, Korevaar D, Bossuyt P. Meta-Analyses of Diagnostic Accuracy in Imaging Journals: Analysis of Pooling Techniques and Their Effect on Summary Estimates of Diagnostic Accuracy. *Radiology*. 2016;281(1):78-85.
6. Willis BH, Quigley M. The assessment of the quality of reporting of meta-analyses in diagnostic research: a systematic review. *BMC Med Res Methodol*. 2011;11:163.
7. Willis BH, Quigley M. Uptake of newer methodological developments and the deployment of meta-analysis in diagnostic test research: a systematic review. *BMC Med Res Methodol*. 2011;11:27.
8. Naaktgeboren CA, van Enst WA, Ochodo EA, et al. Systematic overview finds variation in approaches to investigating and reporting on sources of heterogeneity in systematic reviews of diagnostic studies. *J Clin Epidemiol*. 2014;67(11):1200-1209.
9. Ochodo EA, van Enst WA, Naaktgeboren CA, et al. Incorporating quality assessments of primary studies in the conclusions of diagnostic accuracy reviews: a cross-sectional study. *BMC Med Res Methodol*. 2014;14:33.
10. Naaktgeboren CA, Ochodo EA, Van Enst WA, et al. Assessing variability in results in systematic reviews of diagnostic studies. *BMC Med Res Methodol*. 2016;16(1):6.
11. McGrath T, McInnes M, Langer F, Hong J, Korevaar D, Bossuyt P. Treatment of multiple test readers in diagnostic accuracy systematic reviews-meta-analyses of imaging studies. *European Journal of Radiology*. 2017;93:59-64.
12. McGrath TA, McInnes MDF, van Es N, Leeflang MMG, Korevaar DA, Bossuyt PMM. Overinterpretation of Research Findings: Evidence of "Spin" in Systematic Reviews of Diagnostic Accuracy Studies. *Clin Chem*. 2017.
13. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *J Clin Epidemiol*. 2009;62(10):1006-1012.
14. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *J Clin Epidemiol*. 2009;62(10):e1-34.
15. Macaskill P GC, Deeks JJ, Harbord RM, Takwoingi Y. Chapter 10: Analysing and Presenting Results. In: Deeks JJ, Bossuyt PM, Gatsonis C (editors), Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy The Cochrane Collaboration; 2010.

16. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155(8):529-536.
17. Bossuyt PM, Reitsma JB, Bruns DE, et al. STARD 2015: An Updated List of Essential Items for Reporting Diagnostic Accuracy Studies. *Radiology*. 2015:151516.
18. Cohen JF, Korevaar DA, Altman DG, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *BMJ Open*. 2016;6(11):e012799.
19. Network E. Reporting Guidelines under development. http://www.equator-network.org/library/reporting-guidelines-under-development/#99. Accessed November 28, 2014.
20. Moher D, Schulz KF, Simera I, Altman DG. Guidance for developers of health research reporting guidelines. *PLoS Med*. 2010;7(2):e1000217.
21. McGrath TA, Alabousi M, Skidmore B, et al. Recommendations for reporting of systematic reviews and meta-analyses of diagnostic test accuracy: a systematic review. *Syst Rev*. 2017;6(1):194.
22. Trevelyan E, Robinson N. Delphi methodology in health research: how to do it? *European Journal of Integrative Medicine*. 2015;7:423-428.
23. Boulkedid R, Abdoul H, Loustau M, Sibony O, Alberti C. Using and reporting the Delphi method for selecting healthcare quality indicators: a systematic review. *PLoS One*. 2011;6(6):e20476.
24. Whiting P, Savović J, Higgins JP, et al. ROBIS: A new tool to assess risk of bias in systematic reviews was developed. *J Clin Epidemiol*. 2015.
25. Korevaar DA, Cohen JF, Reitsma JB, et al. Updating standards for reporting diagnostic accuracy: the development of STARD 2015. *Research Integrity and Peer Review*. 2016.
26. Cohen JF, Korevaar DA, Gatsonis CA, et al. STARD for Abstracts: essential items for reporting diagnostic accuracy studies in journal or conference abstracts. *BMJ*. 2017;358:j3751.
27. Korevaar DA, Bossuyt PM, Hooft L. Infrequent and incomplete registration of test accuracy studies: analysis of recent study reports. *BMJ Open*. 2014;4(1):e004596.
28. Korevaar DA, Cohen JF, Spijker R, et al. Reported estimates of diagnostic accuracy in ophthalmology conference abstracts were not associated with full-text publication. *J Clin Epidemiol*. 2016;79:96-103.
29. Deeks J, Bossuyt P, Gatsonis C. *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy* 1.0.0 ed: The Cochrane Collaboration; 2013.
30. van Enst WA, Ochodo E, Scholten RJ, Hooft L, Leeflang MM. Investigation of publication bias in meta-analyses of diagnostic test accuracy: a meta-epidemiological study. *BMC Med Res Methodol*. 2014;14:70.
31. Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *J Clin Epidemiol*. 2005;58(9):882-893.
32. Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Stat Med*. 2001;20(19):2865-2884.

33. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol.* 2005;58(10):982-990.
34. Dinnes J, Mallett S, Hopewell S, Roderick PJ, Deeks JJ. The Moses-Littenberg meta-analytical method generates systematic differences in test accuracy compared to hierarchical meta-analytical models. *J Clin Epidemiol.* 2016;80:77-87.
35. Irwig L, Macaskill P, Glasziou P, Fahey M. Meta-analytic methods for diagnostic test accuracy. *J Clin Epidemiol.* 1995;48(1):119-130; discussion 131-112.
36. Beller EM, Glasziou PP, Altman DG, et al. PRISMA for Abstracts: reporting systematic reviews in journal and conference abstracts. *PLoS Med.* 2013;10(4):e1001419.
37. de Vet HCW EA, Riphagen II, Aertgeerts B, Pewsner D. Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 0.4. 2008.
38. Hong PJ, Korevaar DA, McGrath TA, et al. Reporting of imaging diagnostic accuracy studies with focus on MRI subgroup: Adherence to STARD 2015. *J Magn Reson Imaging.* 2017.
39. Lijmer JG, Mol BW, Heisterkamp S, et al. Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA.* 1999;282(11):1061-1066.
40. Zwinderman AH, Glas AS, Bossuyt PM, Florie J, Bipat S, Stoker J. Statistical models for quantifying diagnostic accuracy with multiple lesions per patient. *Biostatistics.* 2008;9(3):513-522.

Figure Legends

Figure 1: Study flow diagram documenting the Delphi Procedure.

Table 1: Rationale for modification of original PRISMA items for PRISMA-DTA.
Note: final wording for items is provided in Tables 2 and 3.

Item	Reason for modification
1. Title	Specification that the systematic review pertains to 'diagnostic accuracy' in the title is felt to enhance clarity of purpose and allow for easy identification when searching for reviews.
2. Abstract/ Structured summary	As per our study objectives, we aimed to create a specific, essential list of items for DTA systematic reviews to be reported in the abstract. As such, we replaced this item with PRISMA-DTA for abstracts (Table 3).
4. Objectives	The original PRISMA wording (Participant, Intervention, Comparison, Outcome, PICO) was intended for intervention systematic reviews. As such, wording was modified to be more relevant for DTA reviews (e.g. 'Index test' rather than Intervention) [1,29].
6. Eligibility criteria	Language specific to DTA (modifying PICO as described for item 4 above) was added.
8. Search	The primary reason for modification is not specific to DTA, but all contemporary systematic reviews. When the original PRISMA was written, it was likely not feasible to publish all electronic strategies. Present day options for on-line supplemental material and institutional repositories now provide options to report all search strategies. This will enhance transparency, improve replicability and ease systematic review updating.
11. Data items	Data items and relevant definitions with language specific to, and essential, regarding study objectives and risk of bias in DTA reviews (index test, target condition) were modified [16-18,39].
12. Risk of bias in individual studies	Individual studies of DTA may not only be at risk of bias, but there can also be concerns regarding applicability, as highlighted in QUADAS-2. As such, language to reflect this was added [16].
13. Summary measures	1. Summary measures provided in PRISMA (e.g. risk ratios) are specific to systematic reviews of intervention effectiveness. As such, wording was modified to reflect measures relevant to assessing diagnostic accuracy (e.g. sensitivity) [17,18]. 2. The unit of assessment [per-lesion (multiple samples included for every patient as individual data points in a 2x2 table, e.g. multiple liver lesions treated as individual observations) vs. per-patient] can be critical regarding accuracy estimates and generalizability of results due to potential bias introduced from clustering effect in per-lesion analysis [40]; as such this additional requirement relevant to DTA reviews was added.

14. Synthesis of results	1. Measures of consistency (e.g. I^2) considered routine in intervention reviews are not typically applicable in DTA reviews; no consensus regarding alternative statistics presently exists. As such, the more general term, 'variability,' which can reflect multiple strategies to explore variability, was used in place of 'inconsistency' [8,10,15]. 2. Additional specific items particular to DTA reviews were felt to be of sufficient relevance to list as requirements. These include describing definitions of the target condition and test positivity, among others [17,18].
18. Study characteristics	Study characteristics felt to be essential regarding risk of bias and applicability in DTA systematic reviews were listed (e.g. reference standard, clinical setting) [29]. Funding sources was added in order to optimize transparency; industry vs. non-industry funding may be relevant to consider in DTA systematic reviews.
19. Risk of bias within studies	Please see comment for item 12.
20. Results of individual studies	Original wording was specific to systematic reviews of intervention effectiveness: 'summary data for each intervention group'. As such, the wording was revised to reflect results relevant to DTA reviews (e.g. 2x2 data, positivity threshold used) [16-18]. 2x2 data reporting is required to allow readers to evaluate important variables such as the proportion with the target condition, and other accuracy estimates that may not have been specifically addressed in the review (e.g. positive predictive value).
21. Synthesis of results	Language was modified to be more specific for and relevant to DTA systematic reviews e.g. 'describe test accuracy'. In addition, the term 'inconsistency' was replaced with the preferred term for DTA reviews 'variability' as described for item 14 above.
23. Additional analyses	In addition to the original PRISMA wording, we ask that additional information (including potential harms) relevant to DTA systematic reviews be reported [e.g. index test failures (inconclusive, unusable or indeterminate results) and adverse events related to index test administration] [17,18].
24. Summary of evidence	Wording was simplified to only refer to 'main findings' since there is typically only one primary outcome in DTA systematic reviews (diagnostic accuracy). In addition, relevance to key groups was considered to be more appropriate with item 26; as such, it was modified and moved there.
25. Limitations	As per discussion for item 12, wording was modified to reflect concerns regarding 'applicability' in addition to 'risk of bias' [16].

26. Conclusions	As per discussion for item 24, implications for clinical practice were thought to be more appropriate in the conclusions. In addition, language specific to generalizability of findings for DTA reviews (e.g. intended use and clinical role of the index test) was added [17,18] .

Table 2: PRISMA-DTA Checklist. Legend**:** # - Original unmodified PRISMA item; #* - Modified original PRISMA item; D# - New PRISMA-DTA item. *Original PRISMA items 15 and 22 were omitted for reasons outlined in the text.*

Item	Description	
Title/ Abstract		
1*. Title	Identify the report as a systematic review (+/- meta-analysis) of diagnostic test accuracy (DTA) studies.	
2*. Abstract	Abstract: See PRISMA-DTA for abstracts.	
Introduction		
3. Rationale	Describe the rationale for the review in the context of what is already known.	
D1. Clinical role of index test	State the scientific and clinical background, including the intended use and clinical role of the index test, and if applicable, the rationale for minimally acceptable test accuracy (or minimum difference in accuracy for comparative design).	
4*. Objectives	Provide an explicit statement of question(s) being addressed in terms of participants, index test(s), and target condition(s).	
Methods		
5. Protocol and registration	Indicate where the review protocol can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	
6*. Eligibility criteria	Specify study characteristics (participants, setting, index test(s), reference standard(s), target condition(s), and study design) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	
7. Information sources	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	
8*. Search	Present full search strategies for all electronic databases and other sources searched, including any limits used, such that they could be repeated.	
9. Study selection	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	
10. Data collection process	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	
11*. Definitions for data extraction	Provide definitions used in data extraction and classifications of target condition(s), index test(s), reference standard(s) and other characteristics (e.g. study design, clinical setting).	
12*. Risk of bias and applicability	Describe methods used for assessing risk of bias in individual studies and concerns regarding the applicability to the review question.	
13*. Diagnostic accuracy measures	State the principal diagnostic accuracy measure(s) reported (e.g. sensitivity, specificity) and state the unit of assessment (e.g. per-patient, per-lesion).	
14*. Synthesis of results	Describe methods of handling data, combining results of studies and describing variability between studies. This could include, but is not limited to:	
	a) handling of multiple definitions of target condition b) handling of multiple thresholds of test positivity c) handling multiple index test readers	d) handling of indeterminate test results e) grouping and comparing tests f) handling of different reference standards

D2. Meta-analysis	Report the statistical methods used for meta-analyses, if performed.			
16. Additional analyses	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.			
Results				
17. Study selection	Give numbers of studies screened, assessed for eligibility, and included in the review, and, if applicable, included in the meta-analysis, with reasons for exclusions at each stage, ideally with a flow diagram.			
18*. Study characteristics	For each included study provide citations and present key characteristics including:			
	a)	participant characteristics (presentation, prior testing)	e)	index test
	b)	clinical setting	f)	reference standard
	c)	study design	g)	sample size
	d)	target condition definition	h)	funding sources
19*. Risk of bias and applicability	Present evaluation of risk of bias and concerns regarding applicability for each study.			
20*. Results of individual studies	For each analysis in each study (e.g. unique combination of index test, reference standard, and positivity threshold) report 2x2 data (TP, FP, FN, TN) with estimates of diagnostic accuracy and confidence intervals, ideally with a forest or ROC plot.			
21*. Synthesis of results	Describe test accuracy, including variability; if meta-analysis was done, include results and confidence intervals.			
23*. Additional analyses	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression; analysis of index test: failure rates, proportion of inconclusive results, adverse events).			
Discussion/ Conclusions				
24*. Summary	Summarize the main findings including the strength of evidence.			
25*. Limitations	Discuss limitations from included studies (e.g. risk of bias and concerns regarding applicability) and from the review process (e.g. incomplete retrieval of identified research).			
26*. Conclusions	Provide a general interpretation of the results in the context of other evidence. Discuss implications for future research and clinical practice (e.g. the intended use and clinical role of the index test).			
Other				
27. Funding	Describe sources of funding for the systematic review and other support; role of funders for the systematic review.			

Table 3: PRISMA-DTA for Abstracts Checklist.
Legend: #, White = unmodified original PRISMA for Abstracts item**;** #* = modified original PRISMA for Abstracts item; A# = new PRISMA-DTA for Abstracts item.
Original PRISMA for Abstracts item 8 was omitted for reasons outlined in the text.

Item	**Description**
Title/ Purpose	
1*. Title:	Identify the report as a systematic review (+/- meta-analysis) of diagnostic test accuracy (DTA) studies.
2*. Objectives:	The research question including components such as participants, index test(s), and target condition(s).
Methods	
3*. Eligibility criteria:	Study characteristics used as criteria for eligibility.
4. Information sources:	Key databases searched and search dates.
5. Risk of bias and applicability:	Methods of assessing risk of bias and applicability.
A1. Synthesis of results	Methods for data synthesis
Results	
6. Included studies:	Number and type of included studies and participants and relevant characteristics of studies (including reference standard).
7*. Synthesis of results:	Results for analysis of diagnostic accuracy, preferably indicating the number of studies and participants. Describe test accuracy including variability; if meta-analysis was done, include summary results and confidence intervals.
Discussion/ Conclusions	
9*. Strengths and Limitations:	Brief summary of the strength and limitations of the evidence.
10. Interpretation:	General interpretation of the results and important implications.
Other	
11. Funding:	Primary source of funding for the review.
12. Registration:	Registration number and registry name.

Acknowledgements

Conflicts of Interest

The authors have no relevant conflicts of interest to disclose.

Funding

1. Canadian Institute for Health Research (Grant Number 375751)
2. Canadian Agency for Drugs and Technologies in Health (CADTH)
3. STAndards for Reporting of Diagnostic accuracy studies group (STARD)
4. University of Ottawa Department of Radiology Research Stipend Program
5. National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care South West Peninsula

Role of Funders:

None of the funding bodies listed had any role in: design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Access to Data and Data Analysis

Dr. Matthew McInnes had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. There was no substantial data collection or analysis for this work.

Presentation of Statistics and Data

There are no statistical tests provided in this work.